

Compute Working Team Update

State of the Project



Charles Doutriaux, Dan Duffy, Jason Boutte and
Thomas Maxwell, ...

December 6th, 2017



Outline



- Take Home
- Accomplishments
- Things to Come
- Resources

Take Home



- Many Backends Matured
- All compatible via End User API
- COG Integrated Front End
- Ready to be considered as part of installation

Server Side Accomplishments: API



- No changes to Server API!



Server Side Accomplishments: Security



- Authentication added
 - Use OpenID to login

WPS

Login

OpenID Provider

DOE Lawrence Livermore National Laboratory (LLNL)

Login

— Get certificate via

- OAuth
- MyProxyClient

WPS Jobs

Profile Logout

Details Files Processes

Username

cdoutrix

cdoutrix@llnl.gov

https://esgf-node.llnl.gov/esgf-idp/openid/cdoutrix

rd

icated

myproxycient

API Key

pVIC4dq4h0K8sRz4RM5lpyPyFbqykWXgL1Sc8YmHTiR8hLJy7knlTCrcvIG6F11

Update

Regenerate Key

OAuth2

MyProxyClient

Short-Lived Credential Service

Sign in

Sign in with ESGF account

Username: cdoutrix

Password:

Save

Earth System Grid Federation

MyProxyClient Login

Username: cdoutrix

Password:

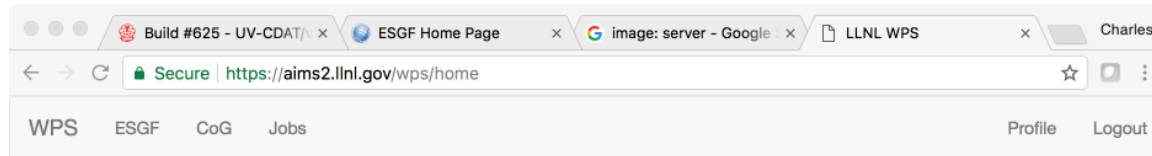
Login



Server Side Accomplishments: Server



- Official Server Implemented:
 - Code: <https://github.com/ESGF/esgf-compute-wps>
 - Docker-based
 - Re-engineered core
 - Celery queues to manage submitted processes
 - Includes LLNL, NASA/EDAS and Ophidia's services
- Deployed at LLNL: <https://aims2.llnl.gov/wps/home>



Welcome to LLNL's CWT WPS server

To get started, login using [OpenID](#).

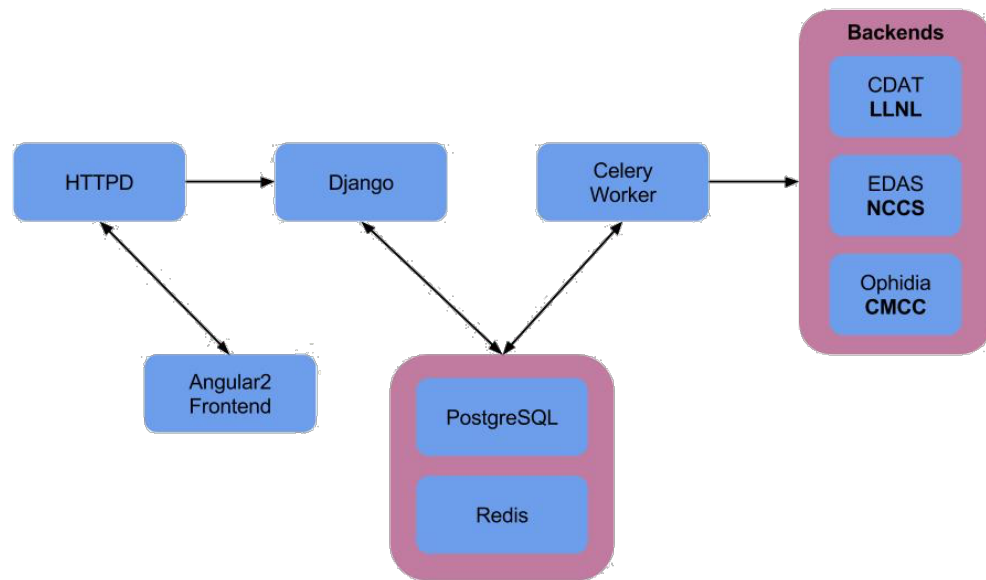
To access ESGF data you will need to retrieve a certificate through OAuth2 or MyProxyClient. These options are found on the bottom of the user [Profile](#) page.

After requesting a certificate, you will find your API key on the user [Profile](#). You can use this to access the ESGF WPS services through then ESGF CWT End-user API which can be install from [Conda](#). Examples of the API can be found [here](#).

Server Side Accomplishments: LLNL

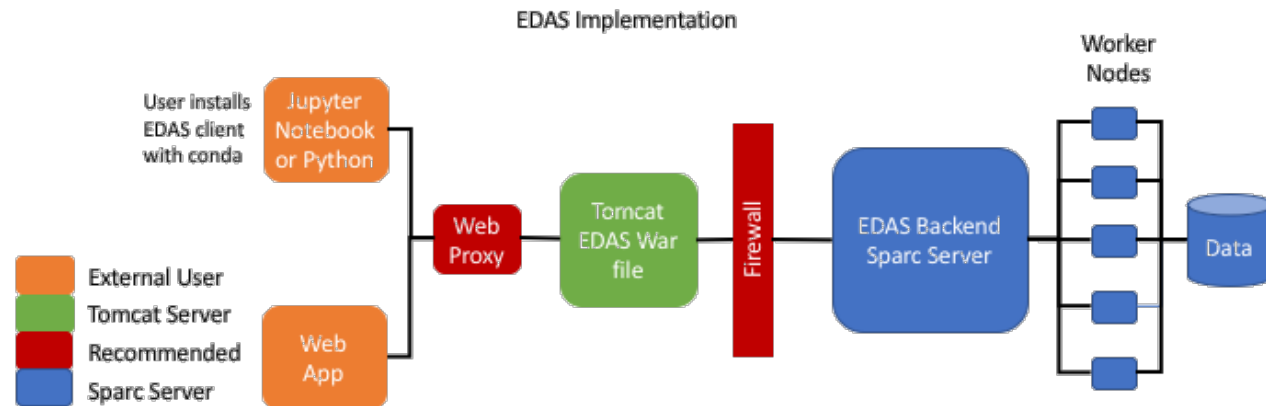
- Services are CDAT-based

- Aggregation
- Average
- Regrid
- Min/Max




Server Side Accomplishments: NASA

- NASA EDAS v. 1.1 deployment on DASS goes public:
 - <https://edas.nccs.nasa.gov/wps/cwt>
- Current Open Source EDAS Distribution:
 - Server: <https://github.com/nasa-nccs-cds/EDAS.git>
 - Web app: <https://github.com/nasa-nccs-cds/CDWPS.git>
 - Client: <https://github.com/ESGF/esgf-compute-api.git>
- Documentation and sample Jupyter Notebooks available at:
 - <https://www.nccs.nasa.gov/services/Analytics>
- List available canonical operations (kernels):
 - <https://edas.nccs.nasa.gov/wps/cwt?request=GetCapabilities>
 - 13 available: emul, ediff, min, emin, max, emax, sum, esum, avg, eavg, rms, erms, ediv
- List available data collections (17 available: MERRA, etc...):
 - <https://edas.nccs.nasa.gov/wps/cwt?request=GetCapabilities&identifier=coll>

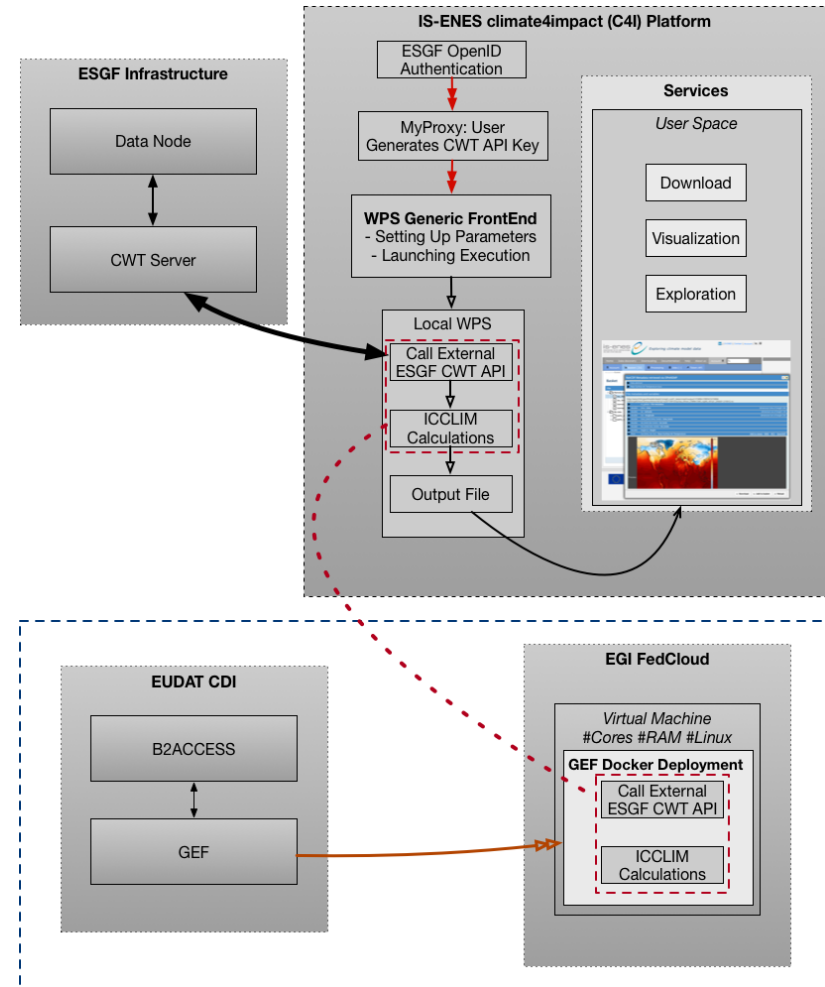


Server Side Accomplishments: Ophidia

- ESGF CWT Plugin for Ophidia implemented
- Properly translates WPS requests into Ophidia requests
- Based on PyOphidia (available on Conda) 
- Leverages the workflow support of Ophidia (through the PyOphidia wsubmit) to implement a WPS process
- Flexible mechanism since each WPS process can be based on several Ophidia operators (workflow)
- Available functionalities: subsetting along any dimension (space and time), maximum & minimum along a specific dimension
- Deployment of the CWT module in the OphidiaLab environment at CMCC

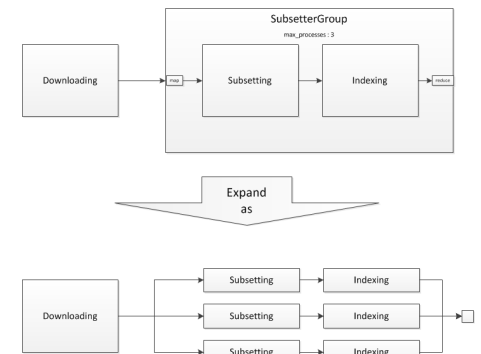
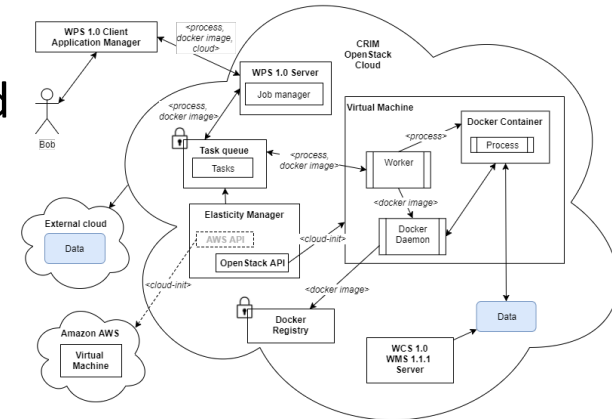
Server Side Accomplishments: CERFACS

- Interfaced CWT and the IS-ENES Climate4impact platform
- Access CWT WPS from within a EUDAT-GEF Execution on EGI



Server Side Accomplishments: CRIM

- Advanced parallel workflow execution
- Prototyped an authorization mechanism for data and processes
- Advanced WPS hybrid cloud execution to integrate with CWT API
- Communicated ESGF uses cases to OGC (security, server-side API, infrastructure, etc.)
- Started work on common test suite
- Contributed to OpenClimateGIS and Birdhouse
- For 2018: more work required to integrate, transfer and harmonize PAVICS and OGC advancements into ESGF software



Server Side Accomplishments: Analytics



- We are already capturing:
 - Files accessed

WPS Jobs Profile Logout

Details Files Processes

Search

Name ▼	Host	Requested	Variable	Last Requested
tas_Amon_CMCC-CM_decadal2005_r1i2p1_200511-201512.nc	aims3.llnl.gov	4	tas	Oct 19, 2017, 3:30:58 PM

Previous Next

- Process launched

WPS Jobs Profile Logout

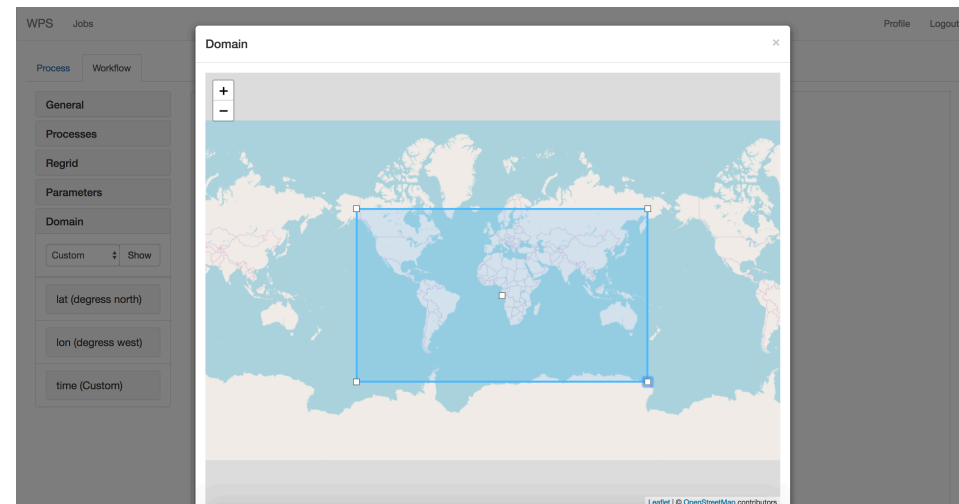
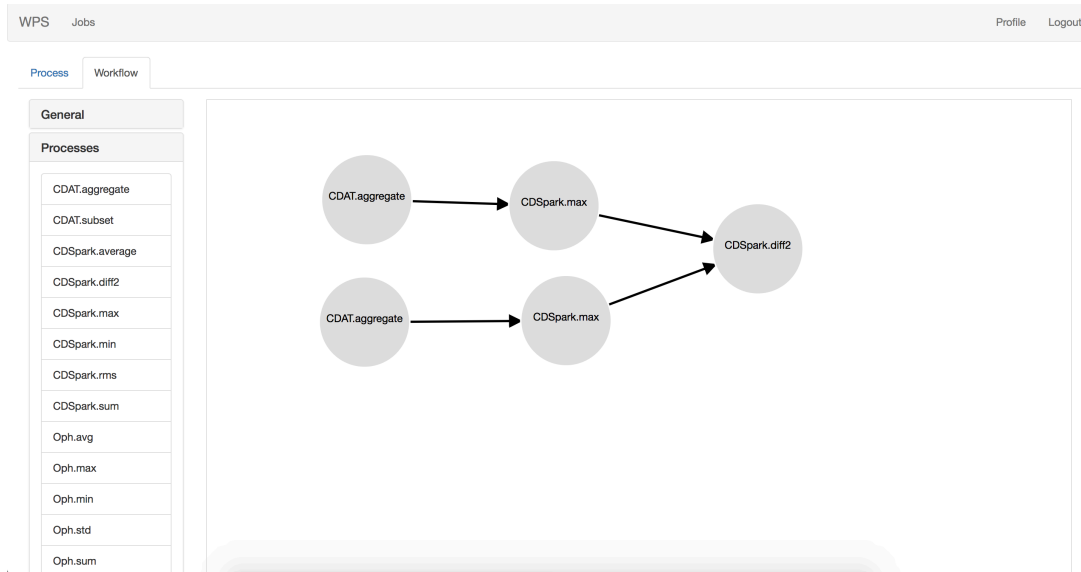
Details Files Processes

Search

Identifier ▼	Backend	Requested	Last Requested
CDAT.subset	Local	4	2017-10-19T22:30:58.084Z

Previous Next

Work In Progress: Workflow and Provenance



COG Integration



Hosted by: < Insert organizational logo(s) here > Powered by **ESGF** and **C&G**

Welcome, Admins | You are a LOCALHOST Node Administrator | Register a New Project | My Profile | Log out

LOCALHOST Home You are at the LOCALHOST node Technical Support

Last Search | My Data Cart (2)

My Data Cart

About Data Carts: You have a Data Cart on every ESGF node you have logged into. This is your Data Cart on the localhost node. The items in this cart will persist until removed.

Number of Items (2) | Return to Last Search

Collective Services for All Selected Datasets: [WGET Script] [LAS Visualization] [Globus Download] [Compute]

When 'List Files' is clicked, or when using WGET or Globus, you may use an optional string to sub-select the filenames:

☒ Select All Datasets

project=CMIP5, model=NASA Goddard Institute for Space Studies, experiment=r1 percent per year CO2, time_frequency=mon, modeling_realm=atmos, ensemble=r11p2, version=20160428
 Description: GISS-E2-R model output prepared for CMIP5 1 percent per year CO2
 Data Node: esgf-node.nas.gov
 Version: 20160428
 Total Number of Files (for all variables): 300
 Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [LAS Visualization] [Compute]

project=CMIP5, model=MIROC5, Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, experiment=r16- or 30-year run initialized in year 2001, time_frequency=mon, modeling_realm=atmos, ensemble=r11p1, version=20120719
 Description: MIROC5 model output prepared for CMIP5 16- or 30-year run initialized in year 2001
 Data Node: esgf-dart.dias.ac.cn
 Version: 20120719
 Total Number of Files (for all variables): 58
 Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [LAS Visualization] [Compute]

ESGF sponsors and partners: DOE Office of Science | (S-ENES) | NASA | NOAA | NCI | NSF C&G version 3.9.7 Earth System C&G sponsors and partners: NOAA | NASA | NSF | DOE Office of Science | (S-ENES)

http://localhost < Insert organizational logo(s) here >

<https://aims2.llnl.gov>

WPS Jobs Profile Logout

Process Workflow

General

Dataset
cmip5.output1.NCC.No ↓

Variable
clw ↓

Process
CDAT.aggregate ↓

Regrid

Parameters

Domain

File Edit View Insert Cell Kernel Widgets Help Trusted Python 2.0

```
proc = wps.get_process('CDAT.subset')
wps.execute(proc, inputs=[tas], domain=d0)

while proc.processing:
    print proc.status
    time.sleep(1)

ProcessStarted Job Started 0 %
```

In [6]: `import cdms2`
`import vcs`
`v = vcs.init()`
`f = cdms2.open(proc.output.uri)`
`v.plot(f['tas'])`

Out[6]:





Things to come

- Full support for OAuth
- Integrated in ESGF release cycle
- Workflows finalized
- Helping other teams' work to be compatible with end-user API
 - Ouranos/Pavics
- Documenting Services
- More Advanced Caching
- Fully distributed
 - Scalability
 - Discovery
- More Services

Resources

- Email: esgf-cwt@llnl.gov
- Webex:
 - First Monday of the month: General Meeting
 - Third Monday of the month: Implementation Meeting
- Documentation
 - Web(requires login)
 - API
- Code (github)
 - Server: <https://github.com/ESGF/esgf-compute-wps>
 - End-user: <https://github.com/ESGF/esgf-compute-apiResources>

Questions Suggested on Agenda

- Define a scalable compute resource (clusters and HPCs) for ESGF data analysis
- Data analytical and visualization capabilities and services
- Performance of model execution
- Advanced networks as easy-to-use community resources (i.e., resource management)
- Provenance and workflow
- Automation of steps for the computational work environment
- Resource management, installation and customer support
- Identify key gaps, identify benefitting communities, and prioritize next steps
- Analysis services when multiple data sets are not co-located (future work)